Optical recognition of converted DNA nucleotides for single-molecule DNA sequencing using nanopore arrays

Ben McNally[1*], Alon Singer[1*], Zhiliang Yu[1], Yingjie Sun[1], Zhiping Weng[2] and Amit Meller[1+]

1. Department of Biomedical Engineering, Boston University, Boston, MA 02215

2. Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA

Supporting Information

Section 1: Circular DNA Conversion (CDC)

Figure SI1 illustrates the conversion of target DNA, a process that we named *Circular DNA Conversion* (CDC) because a circular DNA molecule is formed during each cycle of the conversion. Figure SI1a displays schematically the three steps of CDC, and Figure SI1b displays the results of a single conversion cycle. For proof of principle, we synthesized four single stranded DNA (ssDNA) templates, all 100-nt long and differ only in their 5'-end nucleotide. These templates contain a biotin moiety for immobilization onto streptavidin coated magnetic beads (Invitrogen Dynabeads MyOne Streptavidin C1). This immobilization step enables the quick removal, and replacement, of buffer solutions during the differing stages of the conversion process, with minimal lost of DNA samples. These templates are hybridized to a library of DNA molecules (called probes), each with a double-stranded center portion and two single-stranded overhangs. The double-stranded portion contains the predefined oligonuclotide code that matches the 5'-end nucleotide of the template molecule. Both the 5' template terminal nucleotide and its code are color coded in Figure SI1a: "C" – purple, "A" – yellow, "T" – red and "G" – blue. Only those probes whose 5' overhangs perfectly complement the 5'-end of a template can hybridize with the template. The 3' overhang of the probe hybridizes with the 3'end of the same template

to form a circular molecule. This process is carried out at high salt (100 mM NaCl, 10 mM $MgCl_2$) to promote hybridization. The sample is then washed with a 10 mM TRIS buffer solution, to remove any excess library probes that have not hybridized to the immobilized template molecules.

In the next step of the conversion, the sample is re-suspended in a ligation buffer solution, containing Quick T4 DNA ligase (New England BioLabs), and is used to ligate both ends of the probe with the template (the two locations of ligation are indicated by red circles in Figure SI1a). T4 DNA ligase has been used in other DNA sequencing methods due to its extremely high fidelity compared with other enzymes. After another wash step, to remove the ligation buffer solution, the new circularized sample is re-suspended in a buffer solution containing BseG1 restriction enzyme and a FastDigest buffer (both from Fermantes). The double-stranded portion of the probe contains the recognition site of this restriction enzyme (labeled with an 'R') and positions it to cleave right after the 5'-end nucleotide of the template. This process re-linearizes the circularized molecule in such a way that the predefined code, plus the base that it represents, now reside at the 3' end of the template molecule, and a new base now sits at the 5' end, ready to go through the process of conversion. This process can be repeated as many times as needed, transferring nucleotides from the 5'-end of the template to the 3'-end, interdigitated with the corresponding codes. The conversion of different template molecules does not need to be synchronized, and unproductive hybridization will not lead to error, as long as no ligation and cleavage ensue.

The left panel of Figure SI1b displays a denaturing gel (8 M urea) containing the product after one run of conversion. We see that most of the four different templates were extended by ~50 nts (from 100 to ~150 nts), indicating successful ligation of the template with a probe. To prove that the correct probe was used in each case, we synthesized four types of oligonucleotides as follows: 1) a 16-mer complementary to the "1" bit, with a red fluorophore, 2) a 16-mer complementary to the "0" bit, with a blue fluorophore, 3) a 32-mer complementary to the "10" two-bit sequence, with a green fluorophore, and 4) a 32-mer complementary to "01", with a red fluorophore. A mixture of the first two oligonucleotides was hybridized to each CDC product, and as a control, to all four initial templates. After gel separation, image analysis was carried out

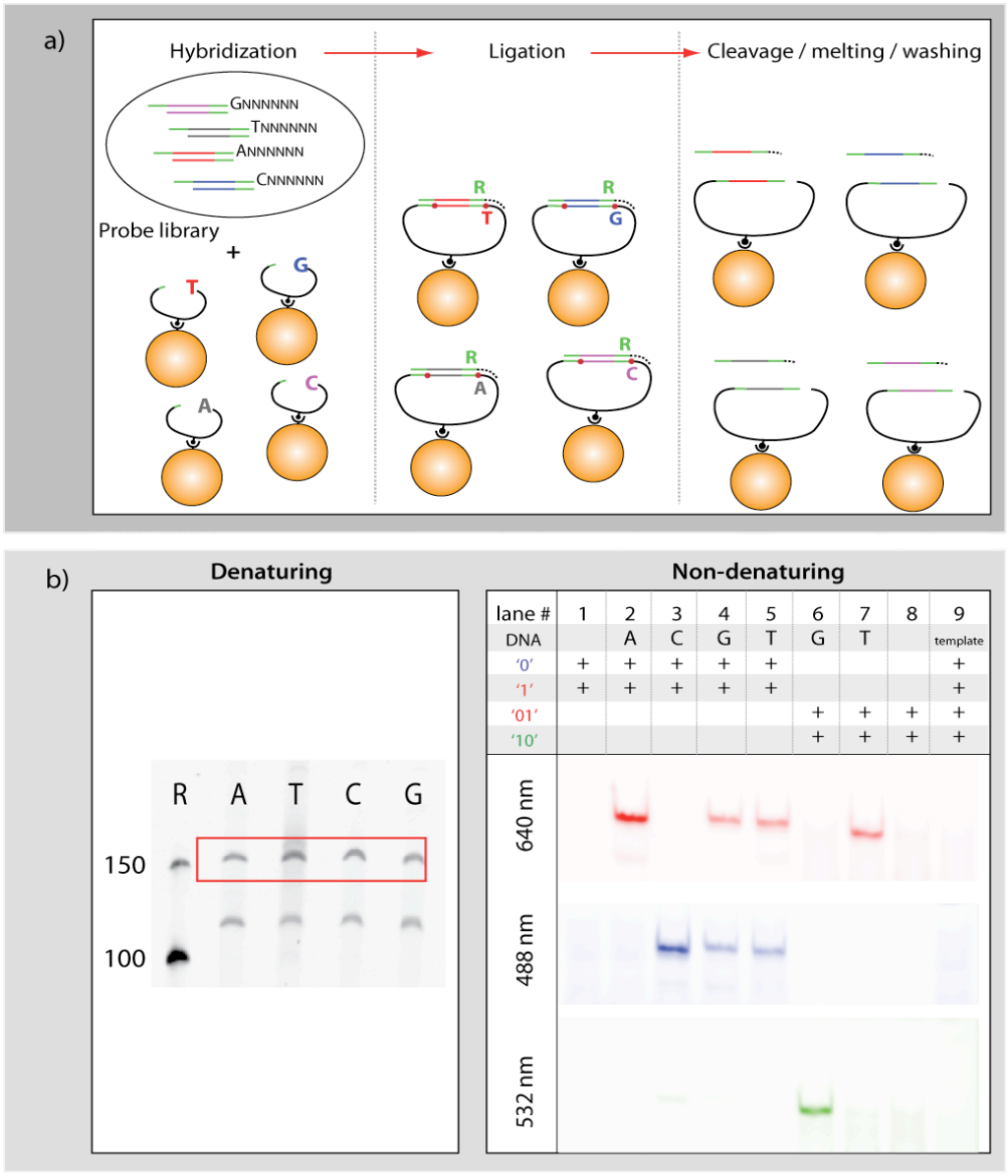

Figure SI1: Circular DNA Conversion (a) Three steps of the CDC. **(b)** Denaturing gel demonstrating successful ligation of probes to all four templates. Lanes A, T, C, and G denote respective 5'-end nucleotides for the four templates, while R is the reference lane containing two ssDNA molecules, 100-nt, and 150-nt in length. **(c)** Using sequence specific fluorescent oligonucleotides, we were able to confirm that the first nucleotides of all four templates were successfully converted and that no by-products result from this process.

using a 3-color laser scanner (BioRad Pharos FX+) and displayed in Figure SI1c. We observe only one red band for the "A" product, only one blue band for the "C" product, coded as "11" and "00" respectively (lane 2 and 3). The other two products, "G" and "T", display both a red and a blue band, as they are coded by "10" and "01" respectively (lane 4 and 5). To distinguish

between the converted “G” and “T”, we hybridize them with the two 32-mer oligonucleotides, as described above. Only “G” displays a band labeled with the green fluorophore, corresponding to the “10” code (lane 6), and only “T” displays a band labeled with the red fluorophore, corresponding to the “01” code (lane 7). Controls show that the templates themselves do not hybridize to any of the labeled oligos, and that the labeled oligos themselves do not show in the gel as they are too short compared with the ~150 nt products (lanes 1, 8 and 9). These results conclusively show that a single CDC cycle produces products with the correct conversion codes.

Section 2: DNA sequences used for proof of principles of Circular DNA Conversion (CDC)

Below are the sequences for the molecular beacons used to verify the identity of the converted products. All the beacon sequences below were synthesized by Eurogentec NA, San Diego:

A. 16-mer complementary to the “1” bit.

5’ – TAAGCGTACGTGCTTA

This sequence has a 5’ amine modification and an ATTO647N (Atto-Tec) dye was conjugated. For nanopore optical readout experiment, the same Oligo was synthesized with a Quencher (BHQ-2, Biosearch Technologies) at the 3’ end.

B. 16mer complementary to the “0” bit:

5’ – CCTGATTCATGTCAGG

This sequence has a 5’ amine modification and an ATTO488 (Atto-Tec) dye was conjugated. For nanopore optical readout experiment, the same Oligo was synthesized with a Quencher (BHQ-2, Biosearch Technologies) at the 3’ end, an ATTO680 (Atto-Tec) dye was conjugated at the 5’ end.

C. 32mer complementary to the “01” sequence:

5’ – CCTGATTCATGTCAGGTAAGCGTACGTGCTTA

This sequence has a 5’ amine modification and an ATTO647N (Atto-Ttec) dye was conjugated.

D. 32mer complementary to the “10” sequence:

5’ – TAAGCGTACGTGCTTACCTGATTCATGTCAGG

This sequence has a 5’ amine modification and a TMR (Invitrogen) dye was conjugated.

Section 3: Bulk Fluorescence Studies

In order to test the efficiency of the quenching process of BHQ-2, we carried out bulk fluorescence experiments. For each fluorophores, we designed two molecules (see insets to Fig SI-2 (a) and (b)). One molecule consisted of a 16mer, containing a fluorescent dye at its 5' end, hybridized to a 66 mer. The second molecule again contained the same 16mer *plus* a second 16mer which contained BHQ-2 quencher at its 3' end. These two 16mers were hybridized to a 66mer. The two 16mer molecules were hybridized such that the fluorescent probe on the 5' end of one was in close proximity to the BHQ-2 quencher on the 3' end of the other. The two fluorophores used were ATTO647N (Atto-Tec) and ATTO680 (Atto-Tec). ATTO647N has a maximum absorption peak at 644 nm and an excitation peak at 669 nm, while ATTO680 has a maximum absorption peak at 680 nm and an excitation peak at 700 nm. For each molecule, we used a spectrofluorometer (JASCO FP-6500) to measure the fluorescence emissions of the complexes. Initially we measured the emissions spectrum of the molecules with the unquenched fluorophores (blue curves in (a) and (b) of Fig SI2). We then measured the emissions spectrum of the molecules with a quencher-fluorophore pair (red curves in (a) and (b) of Fig. SI2). Each

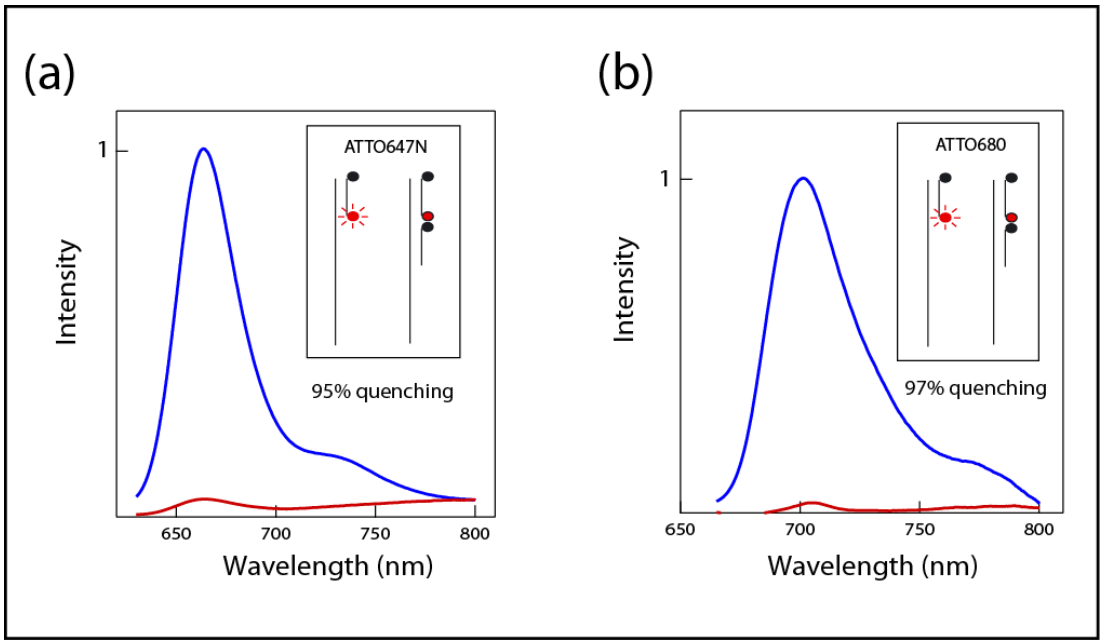

Fig. SI2 (a) Emission spectra for the two complexes containing ATTO647N dye. The blue curve is the measured normalized spectrum for the molecule containing a hybridized ATTO647N beacon, while the red curve is the measured spectrum for the molecule containing both a hybridized ATTO647N beacon as well as a BHQ-2 quencher beacon. The inset to the figure shows schematically the complexes used. **(b)** Emission spectra for the two complexes containing ATTO680 dye. The blue curve is the measured spectrum for the molecule containing a hybridized ATTO680 beacon, while the red curve is the measured spectrum for the molecule containing both a hybridized ATTO680 beacon as well as a BHQ-2 quencher beacon. The inset to the figure shows schematically the complexes used.

experiment contained ~100 nM of hybridized sample. These experiments determined that there is 95-97% quenching occurring for these bulk molecules, as indicated in Fig SI2.